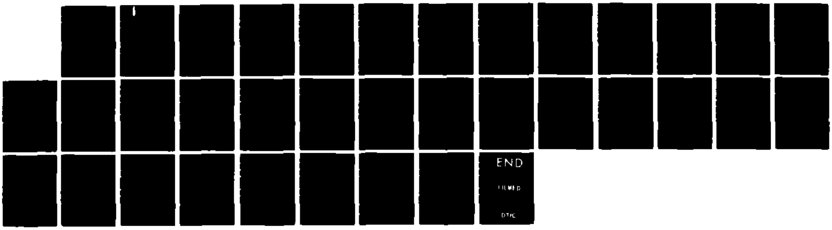
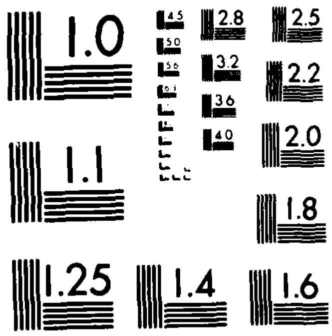
AD-A153 985 PENALTY FUNCTIONS AND DUALITY IN STOCHASTIC PROGRAMMING 1/1
VIA PHI-DIVERGENC..(U) TEXAS UNIV AT AUSTIN CENTER FOR
CYBERNETIC STUDIES A BEN-TAL ET AL. AUG 84 CCS-499
UNCLASSIFIED N00014-82-K-0295 F/G 12/1 NL




1.0
4.5
5.0
5.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

Research Report CCS 499

PENALTY FUNCTIONS AND DUALITY IN STOCHASTIC PROGRAMMING VIA Φ-DIVERGENCE FUNCTIONALS

by

A. Ben-Tal*
M. Teboulle*

CENTER FOR CYBERNETIC STUDIES
The University of Texas
Austin, Texas 78712

Research Report CCS 499

PENALTY FUNCTIONS AND DUALITY IN STOCHASTIC PROGRAMMING VIA Φ-DIVERGENCE FUNCTIONALS

by

A. Ben-Tal*
M. Teboulle*

August, 1984

*The Technion, Israel Institute of Technology, Haifa, Israel

This research was partly supported by ONR Contract N00014-82-K-0295 with the Center for Cybernetic Studies, The University of Texas at Austin and NSF Grant No. ECS-8214081 and the Grant for Promotion of of Research at the Technion. Reproduction in whole or in part is permitted for any purpose of the United States Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
College of Business Administration 5.202
The University of Texas at Austin
Austin, Texas 78712-1177
(512) 471-1821

ABSTRACT

The paper considers stochastically constrained nonlinear programming problems. A penalty type method is suggested as a deterministic surrogate. The penalty is constructed in terms of a "distance" function between random variables, given in term of the ϕ-divergence functional (a generalization of the relative entropy). A duality theory is developed in which a general relation between ϕ-divergence and utility functions is revealed, via the conjugate transform, and a new type of certainty equivalent concept emerges.

Key Words: Stochastic Programming, Penalty Functions, ϕ-divergence, Entropy, Conjugate Duality, Utility Functions, Certainty Equivalent.

1. Introduction

In this paper, we consider mathematical programming problems with stochastic constraints of the form:

$$\text{(SP)} \qquad \inf\{g_0(x): \; g(x,b) \leq 0\}$$

where $x \in \mathbb{R}^n$ is the decision vector, $g_0: \mathbb{R}^n \to \mathbb{R}$, $g: \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$ are given functions and $b \in \mathbb{R}^k$ is a random vector. A new penalty-type decision theoretic approach to treat problem (SP) was introduced recently by Ben-Tal [2]. In this approach, the stochastic program (SP) is replaced by an <u>unconstrained deterministic program</u>:

$$\text{(DP)} \qquad \inf\{g_0(x) + p\,P_E(x)\}$$

where $p > 0$ is a penalty parameter, and P_E is a penalty function for <u>violation of the constraints in the mean</u>, i.e. $P_E(x) = 0$ if $E\,g(x,b) \leq 0$[†] and $P_E(x) > 0$ otherwise.

The special feature of this approach is the choice of the penalty function P_E, which is constructed in terms of the so-called <u>Kullback-Leibler relative entropy functional</u>, (or <u>divergence</u>), widely used in statistical information theory, [9], [10].

If $\mathbb{D}_k$ is the set of all generalized densities f of random vectors $z \in \mathbb{R}^k$ with support T (all absolutely continuous with respect to a common non-negative measure dt), and f_b is a given density in $\mathbb{D}_k$, of the random vector $b \in \mathbb{R}^k$, then the relative entropy between the random vectors z (with

---

[†] E denotes the expectation operator with respect to the random vector b.

density $f \in \mathbb{D}_k$) and b (with density f_b) is

$$I(f,f_b) = \int_T f(t) \log \frac{f(t)}{f_b(t)} dt ; \qquad f \in \mathbb{D}_k .$$

The penalty function P_E is defined in [2] as the following <u>infinite dimensional</u> optimization problem:

$$P_E(x) = \inf_{f \in \mathbb{D}_k} \left\{ \int_T f(t) \log \frac{f(t)}{f_b(t)} dt : \int_T g(x,t) f(t) dt \leq 0 \right\} \tag{1.1}$$

and is accordingly called <u>entropic penalty</u>.

Many attractive properties of the entropic penalty and of the induced deterministic program (DP) are obtained using a fundamental <u>dual representation</u> of P_E derived in [2]:

$$P_E(x) = \sup_{y \geq 0} \{-\log E\, e^{-y^t g(x,b)}\} .$$

In particular, the dual expression of P_E is used to express the deterministic problem (DP) as a saddle function problem, and for the important special case of problems with <u>stochastic right hand side</u>:

(SP-RHS) $\inf\{g_0(x): \; g(x) \geq b\}$

it is shown there, that the <u>primal entropic penalty</u> program (DP) generates a dual problem which consists of maximizing the <u>certainty equivalent</u> of the classical Lagrangian dual function of (SP):

$$h_b(y) := \inf_x \{L_b(x,y) = g_0(x) + y^T g(x,b)\}$$

i.e., the dual problem is:

$$\max_{y \geq 0} u^{-1} Eu(h_b(y))$$

where u is an exponential utility function and u^{-1} is the inverse of u. This interesting dual relationship between the minimization of the classical relative entropy functional and the maximization of expected utility, give rise to the following natural questions:

(1) Does such duality results hold for arbitrary utilities, (not just exponential)?

(2) Assuming (1) holds, what is the corresponding entropy-type functional involved in defining an appropriate penalty function?

(3) How does the new entropy-type penalty relate to utility functions and is it appropriate to treat stochastic programs (SP)?

In this paper, we aim at generalizing and unifying the results derived in [2], and provide satisfactory answers to the above questions. The key to the generalization is the concept of ϕ-divergence, I_ϕ, introduced by Csizar [7]. It includes most of the important entropy type functionals used in mathematical statistics. Its legitimacy as a measure of "distance" between probability distributions as well as some of its basic properties needed in this paper are discussed in Section 2. Adopting this concept here, a generalized penalty function P_ϕ is defined by replacing in (1.1) the classical divergence I (based on the special choice $\phi(t) = t \log t$) with I_ϕ.

In terms of the ϕ-entropic penalty P_ϕ, the stochastic program (SP) is replaced by a deterministic program:

$$(DP)_\phi \qquad \inf_x \{g_0(x) + P_\phi(x)\} \quad .$$

The properties of P_ϕ and its appropriateness in treating stochastic programs

(SP) by its deterministic surrogate $(DP)_\phi$ is discussed in Section 5. A crucial step in studying these properties is the derivation of a simple dual representation of P_ϕ, see Section 3. This representation also enables us to associate in a natural way, the kernel information function ϕ with a <u>utility function</u> u, via the conjugate function ϕ^* of ϕ. In terms of this utility function, we introduce in Section 4 a new type of certainty equivalent concept, possessing for arbitrary utilities many of the properties that the classical certainty equivalent possesses <u>only</u> for exponential utilities. A similar type of such "new certainty equivalent" was first introduced by the authors in [3] from intuitive economic considerations.

In the last section, we treat stochastic right hand side problems, and generalize the results in [2] on the duality between the primal entropic penalty program (DP), and the problem of maximizing the classical certainty equivalent of the Lagrangian dual function $h_b(y)$. It is shown here that the dual problem associated with $(DP)_\phi$, consists of maximizing the <u>new</u> certainty equivalent of $h_b(y)$.

Finally, it is perhaps worthwhile to point out that many other problems which appear in a variety of applications (see [8], [21], [22]) fit the formalism of the ϕ-entropy problem, thus can benefit from the duality framework developed in Section 3. This will be discussed elsewhere in a future paper.

2. <u>The ϕ-Divergence and the Induced ϕ-Entropic Penalty</u>

In this section, we discuss some properties of the ϕ-divergence in terms of which the ϕ-entropic penalty is constructed. Let T be a locally compact Hausdorff space, $\mathcal{F}$ the σ-field of Borel subsets of T, dt a nonnegative regular Borel measure (rBm) on T, and $M(T)$ the linear space of real-valued finite rBm's on T.

Let μ_1 and μ_2 be two probability measures which are assumed absolutely continuous with respect to dt, we denote the density (Radon-Nikodym derivative) of μ_i by: $f_i(t) = \frac{d\mu_i}{dt}$.

Assume here and henceforth in this paper that $\phi\colon \mathbb{R}_+ \to \mathbb{R}$ is a continuous proper convex function, so $\text{dom}\,\phi := \{t\colon\ \phi(t) < \infty\} \neq \emptyset$ and $\phi(t) > -\infty\ \forall t \in \mathbb{R}_+$. The class of such functions will be denoted by Φ.

For a given function $\phi \in \Phi$, the <u>ϕ-divergence</u> of the distributions μ_1 and μ_2 is defined in terms of their densities as:

$$I_\phi(f_1,f_2) := \int_T f_2(t)\,\phi\left(\frac{f_1(t)}{f_2(t)}\right) dt\,. \tag{2.1}$$

The concept of <u>ϕ-divergence</u> (or ϕ-relative entropy) has been introduced by Csiszar[6] as a generalization of many other entropy-type functionals, widely used in statistical information theory (see, e.g. [4],[9],[15],[21]:

Kernel function	The ϕ-divergence	Name/Source
$\phi(t) = t\log t - t + 1$	$I_\phi(f_1,f_2) = \int_T f_1(t)\log\frac{f_1(t)}{f_2(t)}\,dt$	(Kullback-Leibler [10])
$\phi(t) = \frac{1}{2}(1-t)^2$	$I_\phi(f_1,f_2) = \frac{1}{2}\int_T \frac{(f_1(t)-f_2(t))^2}{f_2(t)}\,dt$	(Kagan [9])
$\phi(t) = \frac{1}{\alpha-1}t^\alpha - \frac{t}{\alpha-1} + 1$ $\alpha > 0,\ \alpha \neq 1$	$I_\phi(f_1,f_2) = \frac{1}{\alpha-1}\int_T f_1^\alpha(t) f_2^{\alpha-1}(t)\,dt +$ const.	(α-order divergence [15])
$\phi(t) = (1-\sqrt{t})^2$	$I_\phi(f_1,f_2) = \int_T (\sqrt{f_1}-\sqrt{f_2})^2\,dt$	(Hellinger distance [4])
$\phi(t) = \lvert 1-t\rvert$	$I_\phi(f_1,f_2) = \int_T \lvert f_1(t)-f_2(t)\rvert\,dt$	(Variation distance [21])

TABLE 2.1. Examples of ϕ-divergence

We assume here, and henceforth that $\phi(1) = 0$ and $\lim_{t\to 0^+} \phi(t) = \phi(0)$, $0\phi(\frac{0}{0}) = 0$, $0\phi(\frac{a}{0}) = \lim_{\varepsilon\to 0} \varepsilon\, \phi(\frac{a}{\varepsilon}) = a \lim_{t\to\infty} \frac{\phi(t)}{t}$, $a \in (0,+\infty)$. All the examples in Table 2.1 satisfy these requirements.

The following result follows directly from [7, Lemma 1.1], it explains why I_ϕ can be used as a measure of distance between two random variables.

<u>Proposition 2.1</u> The ϕ-divergence functional (2.1) is well defined and non-negative. It is equal to zero if and only if $f_1 = f_2$ (a.e.) □

The ϕ-divergence also possesses an important convexity property:

<u>Proposition 2.2</u> I_ϕ is convex in each of its arguments.

<u>Proof</u>: The convexity of ϕ is equivalent (for $t > 0$) to that of

$$\phi_o(t) := t\phi(\tfrac{1}{t}) .$$

and

$$I_\phi(f_1,f_2) = I_{\phi_o}(f_2,f_1) = \int_T f_1(t)\phi_o\left(\frac{f_2(t)}{f_1(t)}\right) dt . \tag{2.2}$$

Now, the convexity of I_ϕ in f_1 is obvious, while its convexity in f_2 follows from (2.2).

□

Adopting the concept of ϕ-divergence and following the definition of the entropic penalty P_E given in (1.1), we define now a generalized penalty function $P_\phi(\cdot)$ called <u>ϕ-entropic penalty</u> as

$$\text{(P)} \qquad P_\phi(x) = \inf_{F\in\mathcal{D}_k} \left\{ \int_T pf_b(t)\, \phi\left(\frac{f(t)}{f_b(t)}\right) dt : \int_T g(x,t)f(t)dt \le 0 \right\}$$

Observe that we have built into the definition of P_ϕ a penalty parameter $p > 0$. This parameter enables the decision-maker to control the size of the penalty so as to reflect his subjective attitude towards constraints violations. Note that, by choosing $\phi(t) = t \log t$,[†] one obtains $P_\phi = p \cdot P_E$ where P_E is the usual entropic penalty (see eg. (1.1)). In terms of the ϕ-entropic penalty, a surrogate for the stochastic primal (SP) will be the deterministic primal problem:

$$(DP)_\phi \qquad \inf_{x \in \mathbb{R}^n} \{g_0(x) + P_\phi(x)\} \quad .$$

Properties of the ϕ-entropic penalty and of the induced deterministic program $(DP)_\phi$ will be derived via the duality framework developed in the next section.

3. Duality Theory for the ϕ-Entropic Penalty Problem

Let X and X^* be real vector spaces, and $\langle\cdot,\cdot\rangle$ a bilinear function defined on pairs (x,x^*), $x \in X$, $x^* \in X^*$. Let X and X^* be equipped with locally convex Hausdorff topologies, compatible with the bilinear form, so that every element of one space can be identified with a continuous linear functional on the other. In this case X and X^* are called paired spaces and $\langle\ ,\ \rangle$ is the pairing. For further details, see [4].

Now let X and Y be real vector spaces, $A\colon X \to Y$ a linear operator, $h\colon X \to \mathbb{R}$ a convex function with $\operatorname{dom} h = S$ and $g\colon y \to \mathbb{R}$ a concave function with $\operatorname{dom} g = Q$. Consider the primal problem:

$$(A) \qquad \inf \{h(x)-g(Ax):\ x \in S,\ Ax \in Q\} \quad ,$$

the Fenchel-Rockafellar duality theory [20] associates with (A) the dual problem:

[†]or $\phi(t) = t \log t - t+1$ (as in Table 2.1) which is the normalized form, i.e. $\phi(1) = 0$

(B) $\sup \{g^*(x^*) - h^*(A^* x^*) : \; x^* \in Q^* \;,\; A^* x^* \in S^*\}$

where $A^*: \; Y^* \to X^*$ is the adjoint of A, X^* and Y^* are the spaces paired with X and Y with the pairing $\langle\cdot,\cdot\rangle_x$, $\langle\cdot,\cdot\rangle_y$ respectively, and h^*, g^* are the usual convex and concave conjugates of h and g, i.e.:

$$h^*(\cdot) = \sup_{x \in S} \{\langle x,\cdot\rangle_x - h(x)\}$$

$$g^*(\cdot) = \inf_{y \in Q} \{\langle y,\cdot\rangle_y - g(y)\} \quad .$$

Further, $S^* = \operatorname{dom} h^*$ and $Q^* = \operatorname{dom} g^*$.

The value of the ϕ-entropic penalty at a given point x is obtained as the solution of the infinite dimensional convex optimization problem:

$$\text{(P)} \qquad \inf_{f \in \mathbb{D}_k} \int_T p f_b(t) \, \phi\left(\frac{f(t)}{f_b(t)}\right) dt \qquad\qquad (p > 0)$$

$$\text{subject to} \quad \int_T g_i(x,t) \, f(t) \, dt \leq 0 \qquad i = 1, \ldots, m \, . \tag{3.1}$$

We set problem (P) in the format of the convex program (A) as follows: Consider the linear operator $B: \; M(T) \to \mathbb{R}^m$ given by

$$\mu \to \begin{pmatrix} \int_T g_1(x,t) d\mu \\ \vdots \\ \int_T g_m(x,t) d\mu \end{pmatrix}$$

and the integral functional

$$J(\mu) := p I_\phi(f, f_b) = \begin{cases} \int_T p f_b(t) \, \phi\left(\frac{f(t)}{f_b(t)}\right) dt & \text{if } \mu \text{ is an absolutely continuous 2Bm, and } f = \frac{d\mu}{dt} \\ \infty & \text{otherwise} \end{cases}$$

Let T be the linear function $\mu \to \int d\mu$. Then problem (P) can be written as:

$$\inf\ \{J(\mu):\ \ B\mu \le 0\ ,\ \ T\mu = 1\}\ .$$

By proposition 2.2, J is a convex functional, and it is easily seen that (P) corresponds to the convex program (A) with:

$$S = \mathrm{dom}\ J,\quad X := M(T),\quad Y = \mathbb{R}^{m+2},\quad h(\cdot) := J(\cdot),\quad A := \begin{pmatrix} B \\ T \\ -T \end{pmatrix},$$

$$a := \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},\quad Q := \{z \in \mathbb{R}^{m+2}:\ \ z \le a\}\ \text{ and}$$

$-g(\cdot) := \delta(\cdot|Q)$(= indicator of Q) . Elements $x^* \in Y^*$ are given as: $x^* = (y,\eta_+,\eta_-)$.

Substituting these in the associate dual (B), we have:

Lemma 3.1. The dual problem of (P) is given by

$$\text{(D)}\qquad \sup_{y\in\mathbb{R}^m_+}\ \sup_{\eta\in\mathbb{R}}\ \{\eta - h^*(\eta - y^t g(x,t))\} \tag{3.2}$$

Proof: A simple computation shows that
$$g^*(x^*) = \begin{cases} \eta_+ - \eta_- & \text{if } x^* \le 0 \\ -\infty & \text{otherwise} \end{cases} \tag{3.3}$$

so $Q^* = \{x^* = (y,\eta_+,\eta_-)^t:\ \ x^* \le 0\}$. Further, we note that the adjoints of A and T are respectively:

$$A^*: \mathbb{R}\times\mathbb{R}^2 \to C(T):\quad A^*x^* = B^*y + T^*\eta_+ - T^*\eta_- = y^t g(x,t) + T^*\eta_+ - T^*\eta_- \tag{3.4}$$

$$T^*: \mathbb{R} \to C(T)\qquad T^*s = s\ \text{(A constant function in } C(T)) \tag{3.5}$$

where $C(T)$ is the linear space of continuous function on T, usually identified with the dual space of $M(T)$. Therefore, substituting (3.3), (3.4) and (3.5) in problem (B) with $\eta := \eta_+ - \eta_-$ and replacing y by $-y$, we obtain the desired result. □

In problem (D), the dual objective function is expressed in terms of h^*, which here is the conjugate of the integral functional $J(\cdot)$. The

conjugate of J is computed in the following result.

Lemma 3.2. Let $\phi \in \Phi$, the conjugate $J^*: C(T) \to \mathbb{R}$ of J is given by:

$$J^*(f) = \begin{cases} \int_T pf_b(t)\, \phi^*(\frac{f(t)}{p})\, dt & \text{if } f(t) = \frac{d\mu}{dt}, \ \mu \text{ abs. continuous } (dt) \\ \infty & \text{otherwise} \end{cases} \tag{3.6}$$

Proof: Let $C(T)$ be the space of all continuous functions $x: T \to \mathbb{R}$ with the norm

$$||x|| = \max_{t \in T} |x(t)| .$$

Consider the integral functional $I: C(T) \to \mathbb{R}$ given by:

$$I(x) := \int_T pf_b(t)\, \phi^*(\frac{x(t)}{p})\, dt := \int_T F(t,x)\, dt .$$

The conjugate of I is then by definition:

$$I^*(\mu) = \sup \{ \int_T x d\mu - \int_T F(t,x) dt : \ x \in C(T) \} .$$

Using general results on the computation of I^* (see e.g. [18], Theorem 4 and Corollary 4.A, pp. 452-454) it is easy to verify that for a given $\phi \in \Phi$ and $f_b \in \mathbb{D}_k$, $F(t,x) = pf_b(t)\, \phi^*(\frac{x(t)}{p})$ satisfies the assumptions required there, and so, we have

$$I^*(\mu) = \begin{cases} \int_T F^*(t, \frac{d\mu}{dt}) & \text{if } \mu \in M(T) \text{ is abs. continuous } (dt) \\ \infty & \text{otherwise} \end{cases} \tag{3.7}$$

where $F^*(t,x^*)$ is the conjugate of $F(t,\cdot)$ at x^* (for fixed t).

Here $$F^*(t,x^*) = \sup_{x \in \mathbb{R}} \{ xx^* - pf_b(t)\phi^*(\frac{x}{p}) \} = pf_b(t) \sup_x \left\{ \frac{xx^*}{pf_b(t)} - \phi^*(\frac{x}{p}) \right\}. \tag{3.8}$$

Since $\phi \in \Phi$, hence continuous, we have $\phi = \phi^{**}$ and thus from (3.8):

$$F^*(t,x^*) = \begin{cases} pf_b(t)\ \phi\left(\dfrac{x^*}{f_b(t)}\right) & x^* \geq 0 \\ \infty & \text{otherwise} \end{cases} \qquad (3.9)$$

Setting $x(t) := f(t)$, and combining (3.7)-(3.9), we have obtained

$$I^*(\mu) = J(\mu) \ .$$

Moreover, by the continuity and convexity of I, we have also:

$$J^*(f) = I^{**}(f) = I(f)$$

and $S^* = \operatorname{dom} J^* = C(T)$, and thus (3.6) is proved. □

Combining the results in Lemmas 3.1, 3.2, we have actually proven that the dual problem of (P) is given by:

$$\text{(D)} \qquad \sup_{y \in \mathbb{R}^m_+} \sup_{\eta \in \mathbb{R}} \{\eta - \int_T pf_b(t)\phi^*\left(\frac{\eta - y^t g(x,t)}{p}\right) dt\} \qquad (3.10)$$

a <u>finite dimensional concave</u> program involving only <u>nonnegativity</u> constraints.

Duality results concerning the pair of problems (P) - (D) will now follow.

<u>Theorem 3.1</u>. (a) If (P) is feasible, then inf(P) is attained and

$$\min(P) = \sup(D) \ .$$

Moreover, if there exists a density $f \in \mathbb{D}_k$ satisfying the constraints (3.1) strictly, then sup (D) is attained and

$$\min(P) = \max(D) \ .$$

(b) Under the additional assumption: $\lim_{t\to-\infty} \phi^*(t) < \infty$ then:

sup(D) < ∞ if and only if (P) is feasible .

<u>Proof</u>: (a) The result follows immediately from Rockafellar [20] (Theorems 3,4, pp. 178-179). Indeed, the fact that the dual (D) given in (3.10), has only non-negativity constraints $y \geq 0$, it safisfies the strongest constraint qualification, implying that (D) is <u>stably set</u>, hence the first part of conclusion (a) follows. Also, since $f \in \mathbb{M}_k$ satisfies the constraints (3.1) strictly, (i.e., the familiar Slater regularity condition) then (P) is stably set and thus the second part of (a) is proved.

(b) The implication (P) feasible => sup (D) < ∞ follows trivially from weak duality (without any assumption on the problem (P)). We prove now the reverse implication:

$$(P) \text{ infeasible} \Rightarrow \sup(D) = \infty \quad .$$

The feasible set of (P) is

$$\{B\mu \leq 0, \quad T\mu = 1, \ \mu \text{ nonnegative}\} \quad . \tag{3.11}$$

Using a duality theorem for linear program in vector spaces (see e.g. [13], Theorem 3.13.8, p. 68), it follows that (3.11) is infeasible if and only if the system

$$-B^*y + T^*\eta \leq 0, \quad \eta > 0, \quad y \in \mathbb{R}^m_+ \tag{3.12}$$

is feasible. (B^*, T^* are as defined in (3.4) and (3.5) respectively.) Thus the feasibility of (3.12) implies that:

$$\exists\ \bar{y} \geq 0, \quad \bar{\eta} > 0: \ \bar{\eta} - y^t g\ (x,t) \leq 0, \quad \bar{\eta} > 0 \quad . \tag{3.13}$$

By taking $(\bar{y},\bar{\eta}) \in \mathbb{R}_+^m \times \mathbb{R}_+$ from (3.13), and choosing $y = M\bar{y}$, $\eta = M\bar{\eta}$ with $M > 0$, the dual (D) (see eq. (3.10)) becomes:

$$\sup(D) \geq \sup_{M>0} \{M\bar{\eta} - \int_T pf_b(t)\phi^*(M(\frac{\bar{\eta}-\bar{y}^t g(x,t)}{p}))dt\} \tag{3.14}$$

and since $\lim_{t\to-\infty} \phi^*(t) < \infty$, the sup in (3.14) can be made arbitrary large. □

4. A Representation of the ϕ-Penalty in Terms of a New Certainty Equivalent

Throughout the rest of this paper we deal with the class of function $\phi \in \Phi$ which are strictly convex, essentially smooth(*) (see [17], Section 26), and with $\phi'(1) = 0$. We denote this class by Φ_1.

Recall that a dual representation of the ϕ-entropic penalty is given by (3.10) as:

$$P_\phi(x) = \sup_{y\geq 0} \sup_{\eta\in\mathbb{R}} \{\eta - \int_T pf_b(t)\phi^*(\frac{\eta-y^t g(x,t)}{p})dt\} \quad . \tag{4.1}$$

Let us introduce the <u>utility function</u> u as:

$$u(t) := -\phi^*(-t) \quad . \tag{4.2}$$

Then u is a strictly concave essentially smooth function (see [17], Theorem 26.3) with $u(0) = 0$, $u'(0) = 1$ (this is implied by $\phi(1) = 0$, $\phi'(1) = 0$). Table 4.1 gives the utility functions corresponding to the kernels ϕ given in Table 2.1.

(*) All functions given in Examples 1-4 from Table 2.1 are indeed essentially smooth.

Kernel function $\phi(x)$	Utility function $u(t)$
$x \log x - x + 1$	$1 - e^{-t}$
$\frac{1}{2}(1-x)^2$	$t - \frac{1}{2}t^2$
$(1-\sqrt{x})^2$	$\frac{t}{1+t}$ $(t > -1)$
$\frac{1}{\alpha-1}x^{\alpha} - \frac{\alpha}{\alpha-1}x + 1$	$1 - (1-\frac{t}{\beta})^{\beta}$ $(\frac{1}{\alpha}+\frac{1}{\beta} = 1)$

Table 4.1: The utility functions corresponding to the information-kernel functions.

In terms of u, (4.1) can be written as:

$$P_{\phi}(x) = \sup_{y\geq 0} \sup_{\eta\in\mathbb{R}} \{\eta + pEu(\frac{y^t g(x,b)-\eta}{p})\} \tag{4.3}$$

A little algebra shows that (4.3) can be also written as:

$$P_{\phi}(x) = \sup_{y\geq 0} p\cdot\sup_{\eta\in\mathbb{R}} \{\eta + Eu(\frac{y^t g(x,b)}{p} - \eta)\} \tag{4.4}$$

For a random variable X, let us define the quantity:

$$S_p(X) := p \sup_{\eta\in\mathbb{R}} \{\eta + Eu(\frac{X}{p} - \eta)\} \quad . \tag{4.5}$$

The latter was introduced by the authors in [3] (with $p = 1$), and termed <u>the new certainty equivalent of X</u>.

From (4.4) and (4.5) P_{ϕ} is given by:

$$P_{\phi}(x) = \sup_{y\geq 0} S_p(y^t g(x,b)), \tag{4.6}$$

so the properties of P_{ϕ} are directly related to those of the new-certainty equivalent. The next result summarizes some basic properties of $S_p(X)$.

Lemma 4.1 Let $p > 0$ be fixed. For any random variable X and a constant $w \in \mathbb{R}$:

(a) $S_p(w) = w$

(b) $S_p(X) < E(X)$

(c) $S_p(X+w) = S_p(X) + w$.

Proof: (a) By definition, $S_p(w) = p \sup_{\eta\in\mathbb{R}} \{\eta + u(\frac{w}{p} - \eta)\}$, equating the derivative of the supremand to zero we obtain $u'(\frac{w}{p} - \eta) = 1$, hence since $u'(0) = 1$ and u' is strictly decreasing (as a derivative of the strictly concave function u), the supremum is attained at $\eta = \frac{w}{p}$ and its value is then $S(w) = p \cdot \frac{w}{p} = w$.

(b) Since u is strictly concave, $u(x) < x$ for all $x \neq 0$ hence

$$S_p(X) = p \sup_{\eta\in\mathbb{R}} \{\eta + Eu(\frac{X}{p} - \eta)\} < p \sup_{\eta\in\mathbb{R}} \{\eta + E(\frac{X}{p} - \eta)\} = E(X).$$

(c) By definition:

$$S_p(X+w) = p \sup_{\eta\in\mathbb{R}} \{\eta + Eu(\frac{X+w}{p} - \eta)\} \quad ,$$

hence with $\hat{\eta} = \eta - \frac{w}{p}$, one obtains

$$S_p(X+w) = p \sup_{\hat{\eta}\in\mathbb{R}} \{\hat{\eta} + \frac{w}{p} + Eu(\frac{X}{p} - \eta)\} = w + S_p(X).$$

□

The additiving property given in Lemma 4.1 (c) will be of fundamental importance in deriving the duality results of the next section. Note that property (b) in the lemma corresponds to risk aversion (concave utility).

<u>Example 4.1</u> [Exponential utility]

Let $\phi(x) = x \log x - x + 1$, then $\phi \in \Phi_1$. Its conjugate is $\phi^*(t) = e^t - 1$, so by (4.2), the induced utility is $u(t) = 1 - e^{-t}$. The new certainty equivalent is then

$$S_p(X) = -p\log E\, e^{-X/p} \quad .$$

Here the new certainty equivalent coincides with the classical certainty equivalent corresponding to the utility function $\hat{u}(t) = 1 - e^{-t/p}$, i.e.,

$$S_p(x) = \hat{u}^{-1} E\hat{u}(X).$$

The parameter p is exactly the reciprocal of the Arrow-Pratt risk aversion indicator $(-\hat{u}''/\hat{u}')$, see [14].

<u>Example 4.2</u> [Quadratic utility]

Let $\phi(x) = \frac{1}{2}(x-1)^2$, then $u(t) = t - \frac{t^2}{2}$ and

$$S_p(X) = \mu - \frac{1}{2p}\sigma^2$$

where μ is the mean of X and σ^2 the variance.

Example 4.1 showed that, for exponential utilities, $1/p$ is exactly the classical Arrow-Pratt risk indicator. This role of $1/p$ as a measure of risk, is further explored in the next two results, which incidently provide a generalization of Theorems 1 and 2 derived in Bamberg and Spremann [1] (proved there for the case of exponential utilities only). Lemma 4.3(a) below will be also of particular importance in deriving the duality results of Section 6.

Lemma 4.2 $\lim_{p \to +\infty} S_p(X) = E(X)$ [Risk Neutrality].

Proof: Let $\alpha := \frac{1}{p}$ and define

$$v(\alpha) := \sup_{\eta \in \mathbb{R}} \{\eta + Eu(\alpha X - \eta)\} \tag{4.7}$$

Equating the derivative of the supremand to zero we obtain:

$$Eu'(\alpha X - \eta) = 1 .$$

Since u' is strictly increasing, by the implicit theorem we have

$$v(\alpha) = \eta(\alpha) + Eu(\alpha X - \eta(\alpha)), \tag{4.8}$$

where $\eta(\alpha)$ is the unique solution of

$$Eu'(\alpha X - \eta(\alpha)) = 1 . \tag{4.9}$$

From (4.9), $u'(-\eta(0)) = 1$ and then $\eta(0) = 0$.
In terms of $v(\alpha)$, we have with $\alpha := \frac{1}{p}$

$$S_p(X) = \frac{v(\alpha)}{\alpha} . \tag{4.10}$$

Thus, $\lim_{p \to \infty} S_p(X) = \lim_{\alpha \to 0} \frac{v(\alpha)}{\alpha} = \frac{0}{0}$

and by L'Hopital rule we get:

$$\lim_{p \to \infty} S_p(X) = \lim_{\alpha \to 0} \frac{v'(\alpha)}{1} = \lim_{\alpha \to 0} [\eta'(\alpha) + E((X - \eta'(\alpha))u'(\alpha X - \eta(\alpha))] = E(X).$$

$\square$

Lemma 4.3 Let X be a random variable with infimum support $X_L > -\infty$ and with supremum support X_R. Then,

(a) $\forall \epsilon > 0 \ \lim_{p\to 0} S_p(X) \le X_L + \epsilon$.

(b) If, in addition, u is strictly increasing we have

$$\lim_{p\to 0} S_p(X) = X_L \qquad \text{[Risk averse]}.$$

Proof: (a) A (one dimensional) special case of problem (P) is:

$$(P)_1 \quad \inf_{f\in D_1} \{pI_\phi(f,f_X): \int_{X_L}^{X_R} g(t)f(t)dt \le 0\}$$

Consider now the problem (P_ϵ) obtained from $(P)_1$ with $g(t):=t - X_L - \epsilon$, namely

$$(P_\epsilon) \quad \inf_{f\in D} \{pI_\phi(f,f_X): \int_{X_L}^{X_R} tf(t)dt \le X_L + \epsilon\}$$

The dual of (P_ϵ), obtained from (4.6) is:

$$\sup_{y\ge 0} S_p(y(X - X_L - \epsilon)),$$

and using the additivity of S_p (Lemma 4.1(c)) and (4.7), it becomes:

$$(D\ ) \quad \sup_{y\ge 0} \{S_p(yX) - y(X_L + \epsilon)\} .$$

Problem (P_ϵ) is clearly feasible $\forall \epsilon > 0$, and so from weak duality between the pair $(P_\epsilon) - (D\ )$, we have:

$$(P_\epsilon) \text{ feasible} \Longrightarrow \sup(D_\epsilon) < \infty . \qquad (4.11)$$

Hence,

$$\infty > \sup(D\ ) \ge \lim_{y\to\infty} y\left[\frac{S(yX)}{y} - (X_L + \epsilon)\right]$$

i.e.,

$$\lim_{y\to\infty} \frac{S_p(yX)}{y} \le X_L + \epsilon \quad . \tag{4.12}$$

Now, it is easily verified that $\frac{S_p(yX)}{y} = S_{p/y}(X)$, hence (a) follows from (4.12).

(b) $\alpha > 0$ $\alpha X - \eta \ge \alpha X_L - \eta$ and since u is assumed strictly increasing we have:

$$\eta + Eu(\alpha X - \eta) \ge \eta + u(\alpha X_L - \eta)$$

so

$$S_p(X) \ge p \cdot \sup_{\eta\in\mathbb{R}} \{\eta + u(\frac{X_L}{p} - \eta)\} \quad .$$

The latter supremum is easily computed to be X_L and so

$$S_p(X) \ge X_L .$$

This combined with (a), proves (b). □

5. Properties of P_ϕ and a Min-Max Representation of $(DP)_\phi$

In this section we derive the properties of P_ϕ and discuss its appropriateness in treating the stochastic program (SP), using $(DP)_\phi$.

Theorem 5.1 For any $\phi \in \Phi_1$, the ϕ-entropic penalty function P_ϕ satisfies:

(i)
$$P_\phi(x) = \begin{cases} 0 & \text{if } Eg(x,b) \le 0 \\ \text{positive} & \text{if } Eg(x,b) \not\le 0 \end{cases}$$

(ii) Under the additional assumption (AI): $\lim_{t \to -\infty} \phi^*(t) < \infty$ *

$$P_\phi(x) = \infty \quad \text{if for some } i \quad \underline{g}_i(x) := \inf_{b \in T} g_i(x,b) > 0.$$

Proof: (i) Let $Q(x,y) := S_p(y^t g(x,b))$ then

$$P_\phi(x) = \sup_{y \geq 0} Q(x,y) \geq Q(x,0) = S_p(0) = 0 \tag{5.1}$$

The later equality comes from Lemma 4.1(a).

Now using Lemma 4.1(b) we have:

$$Q(x,y) \leq y^t Eg(x,b)$$

with equality only for $y = 0$ and so

$$P_\phi(x) = \sup_{y \geq 0} Q(x,y) \leq \sup_{y \geq 0} y^t Eg(x,b) .$$

If $Eg(x,b) \leq 0$, the last inequality shows that $P_\phi(x) \leq 0$ which together with (5.1) proves the first part of (i).

Assume now that for some

$$i \in [1,m], \quad Eg_i(x,b) > 0. \tag{5.2}$$

Let $Q_i(x,y_i) := Q(x,0,\ldots,y_i,\ldots,0) = S_p(y_i g_i(x,b))$.

Then, we have $Q_i(x,0) = S_p(0) = 0$.

Moreover,

$$S_p(y_i g_i(x,b)) = \tfrac{1}{p} \cdot \sup_{\eta \in \mathbb{R}} \{\eta + Eu(\frac{y_i g_i(x,b)}{p} - \eta)\} . \tag{5.3}$$

Since $\phi \in \Phi_1$, then the function $\psi_x(y_i,\eta) := Eu'(\frac{y_i}{p} g_i(x,b) - \eta)$ is continuously differentiable on $\mathbb{R}_+ \times \mathbb{R}$ and $\frac{d}{d\eta} \psi_x(y,\eta) = -Eu''(\frac{y_i}{p} g_i(x,b) - \eta) > 0$.

---

* This assumption holds for Examples 1 (with $\alpha < 1$), 2 and 4 in Table 4.1

Hence, by the Implicit Function Theorem, there exists a unique solution $\eta = \eta(y_i)$ to the equation $\psi_x(y_i,\eta(y_i)) = 1$, and $\eta(0) = 0$. By the definition of $S(y_i g_i(x,b))$ in (5.3), as an unconstrained concave optimization problem, an explicit expression of it is obtained by equating the derivative of the supremand to zero, and therefore:

$$Q_i(x,y_i) = S_p(y_i g_i(x,b) = p(\eta(y_i) + Eu(\frac{y_i g_i(x,b)}{p} - \eta(y_i))) \tag{5.4}$$

where $\eta(y_i)$ is the unique solution of

$$\psi_x(y_i,\eta) = 1 \ . \tag{5.5}$$

Now, an easy computation shows that:

$$\frac{d}{dy_i} Q_i(x,y_i)\Big]_{y_i=0} = Eg_i(x,b)\cdot u'(-\eta(0)),$$

but $\eta(0) = 0$ and $u'(0) = 1$, so we have under assumption (5.2):

$$\frac{d}{dy_i} Q_i(x,y_i)\Big]_{y_i=0} = Eg_i(x,b) > 0.$$

Therefore, there exist $\hat{y}_i > 0$ (close enough to zero) such that

$$Q_i(x,\hat{y}_i) > Q_i(x,0) = 0 \tag{5.6}$$

Noting that

$$P_\phi(x) = \sup_{0\leq y \in \mathbb{R}^n} Q(x,y) \geq \sup_{0\leq y_k \in \mathbb{R}} Q_i(x,y_i) \geq Q_i(x,\hat{y}_i) > 0,$$

and this proves the second part of (i).

(ii) Let $g_i(x) > 0$ for some i. The latter means that problem (P) (see (3.1)) is infeasible, hence under assumption (A1), invoking Theorem 3.1(b), this implies that $P_\phi(x) = \infty$. □

The first part of the theorem demonstrates that $P_\phi(x)$ is a penalty function for violation of the constraints in the mean. The second part shows that P_ϕ has the desirable property of excluding solutions <u>which are not feasible</u> in (SP), <u>for any realization of b</u>, i.e., for x which is infeasible with probability 1, $P_\phi(x) = \infty$.

Therefore, a suitable deterministic surrogate problem for (SP) is

$$(DP)_\phi \quad \inf_{x \in \mathbb{R}^n} \{g_o(x) + P_\phi(x)\} \quad .$$

From the additivity of he new certainty equivalent (Lemma 4.1(c))

$$\inf_x \{g_o(x) + P_\phi(x)\} = \inf_x \sup_{y \geq 0} S_p(g_o(x) + y^t g(x,b))$$

hence $(DP)_\phi$ can be written as a <u>minimax problem</u>:

$$(DP)_\phi \quad \inf_x \sup_{y \geq 0} S_p(L_b(x,y)) \tag{5.7}$$

where $L_b(x,y) = g_o(x) + y^t g(x,b)$ is the classical Lagrangian corresponding the the original problem (SP). This result generalizes Theorem 2 of [2].

We close this section by giving an explicit expression of P_ϕ for the familiar chance constraints problem [6].

<u>Example 5.1</u>: Consider the well known chance constrained program

$$(CC) \quad \inf \{g_o(x): \Pr\{g(x) \geq b\} \leq \alpha\}^* \tag{5.8}$$

which is a special case of the deterministic program :

$$\inf \{g_o(x): Eg(x,b) \leq 0\} \tag{5.9}$$

---

* For simplicity we will treat here only the case of a single constraint, i.e., $g: R^n \to R$.

by choosing:

$$g(x,b) = \begin{cases} 1-\alpha & \text{if } g(x) \geq b \\ -\alpha & \text{if } g(x) < b \end{cases} \qquad \alpha \in (0,1) \tag{5.10}$$

Using the dual representation of P_ϕ given in (4.4) we have:

$$P_\phi(x) = \sup_{0\leq y\in\mathbb{R}} p \sup_{\eta\in\mathbb{R}} \{\eta + \int_{-\infty}^{\infty} f_b(t)u(\frac{y}{p}\, g(x,t) - \eta)dt\} \tag{5.11}$$

Recalling from (5.10) that here:

$$g(x,t) = \begin{cases} 1-\alpha & \text{if } g(x) \geq t \\ -\alpha & \text{otherwise} \end{cases} \qquad \alpha \in (0,1)$$

we get from (5.11) in term of the cummulative distribution function $F(\cdot)$ of b: (denoting $F:=F(g(x))$):

$$P_\phi(x) = \sup_{y\geq 0} p \sup_{\eta\in\mathbb{R}} \{\eta + Fu(\frac{y(1-\alpha)}{p} - \eta) + (1-F)u(-\frac{\alpha y}{p} - \eta)\} \quad . \tag{5.12}$$

Let us define $\hat{\eta} := -\frac{\alpha y}{p} - \eta$, then (5.12) becomes

$$P_\phi(x) = p \sup_{\hat{\eta}\in\mathbb{R}} \{-\hat{\eta} + (1-F)u(\hat{\eta}) + \sup_{y\geq 0} \{-\frac{\alpha y}{p} + Fu(\hat{\eta} + \frac{y}{p})\}\tag{5.13}$$

The inner supremum in (5.13) is computed first; by simple calculus, the maximizing y is y^* given by:

$$y_i^* = \begin{cases} p(u')^{-1}(\frac{\alpha}{F}) - \hat{\eta} & \text{if } \frac{\alpha}{F} \geq u'(\hat{\eta}) \\ 0 & \text{otherwise.} \end{cases}$$

(The existence of $(u')^{-1}$ is guaranteed, since u' is a derivative of a strictly concave function.)

Substituting y^* in (5.13) yields:

$$P_\phi(x) = \begin{cases} p \sup\limits_{\hat{\eta}\in\mathbb{R}} \{(1-F)u(\hat{\eta}) - \hat{\eta}(1-\alpha)\} + \psi(F) & \text{if } \frac{\alpha}{F} \geq u'(\hat{\eta}) \\ 0 & \text{otherwise} \end{cases} \tag{5.14}$$

where $\psi(t) := -t\cdot\{\frac{\alpha}{t}(u')^{-1}(\frac{\alpha}{t}) - u((u')^{-1}(\frac{\alpha}{t}))\}$. (5.15)

The latter expression can be simplified by observing that

$$u^*(x^*) = \inf_x \{xx^* - u(x)\} = x^*(u')^{-1}(x^*) - u((u')^{-1}(x^*)) \tag{5.16}$$

(i.e., the Legendre Transform of u), but by (4.2) we know that

$$u(x) = -\phi^*(-x), \text{ hence } u^*(x^*) = -\phi(x^*)$$

using the latter in (5.16) at the point $x^* = \frac{\alpha}{t}$, (5.15) reduces to the simple expression:

$$\psi(t) = t\phi(\frac{\alpha}{t}) \quad . \tag{5.17}$$

It remains to compute

$$\sup_{\hat{\eta}\in\mathbb{R}} \{(1-F)u(\hat{\eta}) - \hat{\eta}(1-\alpha)\} \ . \tag{5.18}$$

Equating the derivative of the supremand to zero, we get the optimal $\hat{\eta}^*$ from:

$$u'(\hat{\eta}^*) = \frac{1-\alpha}{1-F}$$

which by (5.14) must satisfy $\frac{\alpha}{F} \geq u'(\eta^*) = \frac{1-\alpha}{1-F}$; then substituting $\hat{\eta}^*$ in (5.18), after some algebra, we finally get from (5.14):

$$P_\phi(x) = \begin{cases} p\cdot\{F(g(x))\phi(\frac{\alpha}{F(g(x))}) + (1-F(g(x))\phi(\frac{1-\alpha}{1-F(g(x))})\} & \text{if } F(g(x)) \geq \alpha \\ 0 & \text{if } F(g(x)) < \alpha \\ & \text{i.e. } \Pr\{g(x) \geq b\} \leq \alpha \end{cases} \tag{5.19}$$

Note that the function $h(t) := t\phi(\frac{\alpha}{t}) + (1-t)\phi(\frac{1-\alpha}{1-t})$ in term of which P_ϕ is expressed, is convex and increasing $0 < \alpha \leq t < 1$ and $h(\alpha) = 0$, so by this and (5.19), $P_\phi(x) = h(\max\{\alpha, F(g(x))\})$. It follows that if $F(g(x))$ is convex so is P_ϕ. (Compare these results with [2].)

6. The Dual Problem of $(DP)_\phi$ for Right-Handside Programs

In this section we treat the special case of the general problem (SP):

$$\text{(SP)} \quad \inf\{g_o(x) : g_i(x) \geq b_i \quad i = 1, \dots, m\}$$

which is obtained from (SP) by setting $g(x,b) := b - g(x)$.

The penalty function P_ϕ is given here by:

$$P_\phi(x) = \sup_{y \geq 0} S_p(y^t(b - g(x))) .$$

However, by the additivity of S_p, since $y^t g(x)$ is not random, we can write P_ϕ as:

$$P_\phi(x) = \sup_{y \geq 0} \{w(y) - y^t g(x)\}$$

where

$$w(y) := p \sup_{\eta \in \mathbb{R}} \{\eta + Eu(\frac{y^t b}{p} - \eta)\} \quad . \tag{6.1}$$

The corresponding deterministic primal $(DP_\phi - \text{RHS})$: $\inf_x \{g_o + P_\phi(x)\}$ is then:

$$(DP_\phi - \text{RHS}) \quad \inf_x \sup_{y \geq 0} K(x,y)$$

with

$$K(x,y) := g_o(x) + w(y) - y^t g(x) \tag{6.2}$$

Assume now, that $g_o(x)$ is convex, and that $\{g_i(x)\}_{i=1}^m$ are concave

functions, so (SP - RHS) is a convex program. Also $P_\phi(x)$ becomes convex and so $(DP_\phi$ - RHS) is a convex program.

We define the dual problem corresponding to $(DP_\phi$ - RHS) by:

$$(DD_\phi - RHS): \quad \sup_{y \geq 0} \inf_{x} K(x,y) \quad .$$

The main result of this section is a strong duality relation between the pair $(DP_\phi$ - RHS) and $(DD_\phi$ - RHS).

Theorem 6.1 Let (SP-RHS) be a convex stochastic program and consider the corresponding deterministic program $(DP_\phi$ - RHS).

If the following condition holds:

$$\exists\ \hat{x} \in \mathbb{R} \quad \text{such that} \quad g_i(\hat{x}) > \underline{b}_i \qquad \forall\ i \in I \tag{6.3}$$

where $\underline{b}_i$ denote the infinum support of b_i.
Then:

$$\inf\ (DP_\phi - RHS) = \max(DD_\phi - RHS)\ . \tag{6.4}$$

Proof: Since $g_o(x)$ is convex and $\{g_i(x)\}_{i=1}^m$ are concave then $K(\cdot,y)$ given in (6.2) is convex for every $y \geq 0$.
Now the function $w(\cdot)$ given in (6.1) can be rewritten as:

$$w(y) = - \inf_{\eta \in \mathbb{R}} F(\eta,y),$$

where $F\colon \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$ is defined by $F(\eta,y) = -p\eta - pEu(\frac{y^t b}{p} - \eta)$; since u is concave it follows from [19], Theorem 1 that $F(\eta,y)$ is (jointly) convex, hence $w(y)$ is concave and therefore $K(x,\cdot)$ is concave.

By a result in [16], a sufficient condition for the validity of (6.4) for a general convex-concave saddle function $K(x,y)$ is:

$$\exists\ y_o \geq 0 \text{ such that } y_o^t \nabla y K(x,y) \geq 0 \quad (x \in \mathbb{R}^n,\ y > 0) \tag{6.5}$$

With analogous proof to the one given in Theorem 5.1 (eqs. (5.4)-(5.5), $w(y)$ is expressed as

$$w(y) = p(\eta(y) + Eu(\frac{y^t b}{p} - \eta(y))$$

where $\eta(y)$ is obtained from $Eu'(\frac{y^t b}{p} - \eta(y)) = 1$.

Thus, $\nabla w(y) = E(bu'(\frac{y^t b}{p} - \eta))$, hence (6.5) is here:

$$\nexists\ y_o \geq 0 \text{ such that } y^t\{E(bu'(\frac{y^t b}{p} - \eta(y)) - g(x))\} \geq 0$$

$$\forall x \in \mathbb{R}^m,\ y > 0.$$

The later is certainly satisfied if:

$$\exists\ \hat{x}, \hat{y} > 0 \text{ such that } \nabla w(y) = E(bu'(\frac{y^t b}{p} - \eta(y)) < g(\hat{x}) \tag{6.6}$$

To show that condition (6.3) implies (6.6) it suffices to prove that:

$$\inf_{y \geq 0} \frac{\partial}{\partial y_i} w(y) \leq \underline{b}_i \qquad \forall i = 1,\dots m \quad . \qquad (6.7$$

For all $i \in [1,m]$, let

$$w_i(y_i) := w(0,\dots,y_i,\dots 0) = p \sup_{\eta \in \mathbb{R}} \{\eta + Eu(\frac{b_i y_i}{p} - \eta)$$

Noting that

$$\inf_{0 \leq y \in \mathbb{R}^m} \frac{\partial}{\partial y_i} w(y) \leq \inf_{0 \leq y_i \in \mathbb{R}} \frac{\partial}{\partial y_i} w_i(y_i) \quad \forall i$$

to prove (6.7) it suffices to prove that

$$\inf_{0 \leq y_i} w_i'(y_i) \leq \underline{b}_i \tag{6.8}$$

Now $w_i(y)$ is concave and $w_i(0) = 0$, hence by the gradient inequality

$$0 = w_i(0) \leq w_i(y_i) - y_i w_i'(y_i)$$

and thus

$$\lim_{y_i \to +\infty} w_i'(y_i) \le \lim_{y_i \to \infty} \frac{w(y_i)}{y_i} \tag{6.9}$$

But w_i' is a derivative of a strictly concave function and thus is strictly decreasing hence $\inf\limits_{y_i \geq 0} w_i'(y_i) = \lim\limits_{y_i \to \infty} w_i'(y_i)$.

Moreover, using (4.7) we have the relation:

$$w_i(y_i) = y_i \frac{v(\alpha)}{\alpha}$$

with $\alpha := \frac{y_i}{p}$ $(p > 0)$, then from (6.9) and Lemma 4.3(a) we get the desired result (6.8).

□

The dual problem (DD_ϕ - RHS) is given by

$$\sup_{y \geq 0} \inf_x K(x,y) .$$

To get a full meaning of this dual we fist prove:

<u>Lemma 6.1</u> $\inf\limits_x S_p(L_b(x,b)) = S_p(\inf\limits_x L_b(x,y))$.

<u>Proof</u>:

$$\begin{aligned} S_p(\inf_x L_b(x,y)) &= S_p(y^t b + \inf_x (g_o(x) - y^t g)x))) \\ &= S_p(y^t b) + \inf_x \{g_o(x) - y^t g(x)\} \quad \text{[by Lemma 4.1(c)]} \\ &= \inf_x \{S_p(y^t b) + g_o(x) - y^t g(x)\} \\ &= \inf_x \{S_p(y^t b + g_o(x) - y^t g(x)\} \quad \text{[by Lemma 4.1(c)]} \\ &= \inf_x \{S_p(L_b(x,y))\}. \end{aligned}$$

□

Now $K(x,y)$ given in (6.2) can be written, using again the additivity property of S_p (Lemma 4.1(c)) as:

$$K(x,y) = S_p(L_b(x,y)) \ .$$

Hence, by Lemma 6.1, the dual problem $(DD_\phi - RHS)$ is

$$(DD_\phi - RHS) \quad \sup_{y \geq 0} S_p(\inf_x L_b(x,y)) \ .$$

Thus we have shown that while in the deterministic case, the Lagrangian dual of (SP) is the concave program: $\sup_{y \geq 0} \inf_x L_b(x,y)$, in the stochastic case, the dual program $(DD_\phi - RHS)$ consists of maximizing the new certainty of the Lagrangian dual function.

This result generalizes, to arbitrary utilities, a result in [2, Theorem 4], proved for the exponential utility.

REFERENCES

1. Bamberg, G. and Spremann, K., "Implications of Constant Risk Aversion", Zeit. fur Operations Research, 25 (1982), pp. 205-222.

2. Ben-Tal, A., "The Entropic Penalty Approach to Stochastic Programming", to appear in Math. of Oper. Res.

3. Ben-Tal, A., and Teboulle, M., "Expected Utility, Penalty Functions, and Duality in Stochastic Non-Linear Programming". Operations Research, Statistics and Economics, Mimeograph Series, No. 366, 1984, Technion.

4. Beran, R., "Minimum Hellinger Distance Estimates for Parametric Models", The Annals of Statistics, 5 (1977), pp. 445-463.

5. Bourbaki, N., Espaces Vectoriels Topologiques, Hermann et Cie., Paris, 1955.

6. Charnes, A. and Cooper, W.W., "Chance Constrained Programming", Manag. Sci., 5, (1959), pp. 73-79.

7. Csiszar, I., "Information-Type Measures of Difference of Probability Distributions and Indirect Observations". Studia Sci. Mat. Hungar., 2, (1967), pp. 299-318.

8. Ferentinos, K. and Papaioannov, T., "New Parametric Measures of Information", Information and Control, 51, (1981), pp. 193-208.

9. Kagan, A.M., "On the Theory of Fisher's Amount of Information", Sov. Math., Dokl, 4, (1963), pp. 991-993.

10. Kullback, S., Leibler, R.A., "On Information and Sufficiency", Ann. Math. Stat., 22, (1951), pp. 79-86.

11. Kullback, S., Information Theory and Statistics, John Wiley and Sons, Inc., New York, (1959).

12. Laurent, P.M., Approximation et Optimisation, Hermann et Cie, Paris, 1972.

13. Ponstein, J., Approaches to the Theory of Optimization, Cambridge University Press, 1980.

14. Pratt, J.W., "Risk Aversion in the Small and in the Large", Econometrica, 32, (1964), pp. 122-136.

15. Renyi, A., "On Measures of Entropy and Information", in Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 1961, Vol. 1, pp. 547-561.

16. Rockafellar, R.T., "Minimax Theorems and Conjugate Saddle Functions", Math. Scand., 14, (1964), pp. 151-173.

17. Rockafellar, R.T., Convex Analysis, Princeton University Press, Princeton, N.J., 1970.

18. Rockafellar, R.T., "Integrals which are Convex Functions, II", Pacific J. Math., 39, (1971), pp. 439-469.

19. Rockafellar, R.T., Conjugate Duality and Optimization. Regional Conference Series in Applied Mathematics, No. 16, SIAM, 1974.

20. Rockafellar, R.T., "Duality and Stability in Extremum Problems Involving Convex Functions", Pacific J. Math., 21, (1976), pp. 163-186.

21. Vajda, I., "On the f-Divergence and Singularity of Probability Measures", Periodica Mathematica Hungarica, 2, (1972), pp. 223-234.

22. Ziv, J., and Zakai, M., "On Functionals Satisfying a Data-Processing Theorem", IEEE Trans. on Information Theory, IT19, (1973), pp. 275-283.

END

FILMED

6-85

DTIC